MALRIS: Malicious Hardware in RIS-Assisted Wireless Communications

Danish Mehmood Mughal, Daniyal Munir, Qazi Arbab Ahmed, Hans D. Schotten, Thorsten Jungeblut, Sang-Hyo Kim, and Min Young Chung

Abstract

Reconfigurable intelligent surfaces (RIS) enhance wireless communication by dynamically shaping the propagation environment, but their integration introduces hardware-level security risks. This paper presents the concept of Malicious RIS (MALRIS), where compromised components behave adversarially, even under passive operation. The focus of this work is on practical threats such as manufacturing time tampering, malicious firmware, and partial element control. Two representative attacks, power-splitting and element-splitting, are modeled to assess their impact. Simulations in a RIS-assisted system reveal that even a limited hardware compromise can significantly degrade performance metrics such as bit error rate, throughput, and secrecy metrics. By exposing this overlooked threat surface, this work aims to promote awareness and support secure, trustworthy RIS deployment in future wireless networks.

Index Terms

Reconfigurable Intelligent Surfaces (RIS), Hardware Security, Hardware Trojans, 6G, Malicious RIS.

This is the author's version of the work accepted for presentation at IEEE CSCN 2025.

I. Introduction

Reconfigurable Intelligent Surfaces (RIS) are emerging as a cornerstone technology in the evolution of wireless networks, particularly in the context of 6G [1]. By enabling the dynamic manipulation of the wireless propagation environment, RIS can significantly enhance signal strength, coverage, and energy efficiency through the passive reflection and phase control of electromagnetic waves [2]. These capabilities position RIS as a key enabler of an intelligent, programmable, and highly efficient communication infrastructure [3].

In recent years, RIS has gained unprecedented attention from academia and industry (see [4] and references therein) for its potential in future wireless networks. By enabling programmable manipulation of the wireless environment, RIS extends coverage in non-line-of-sight and shadowed areas, overcoming obstructions and dead zones [5]. Through intelligent phase control, it enhances spectral efficiency, improving SINR and data rates [4]. RIS operates passively without amplifying or generating new signals, offering greater energy efficiency than active repeaters or relays [6]. It also mitigates interference and improves link reliability by dynamically steering beams based on real-time channel conditions [7]. Additionally, RIS supports physical layer security by directing energy to intended receivers while limiting leakage to eavesdroppers [8]. These features reduce latency, enhance reliability, and enable cost-effective network scaling, meeting key 6G performance targets.

Despite these promising capabilities, most RIS research has focused on performance gains via signal processing, channel estimation, and optimization [4], assuming RISs are passive and trusted. However, as RISs are increasingly deployed in exposed environments, their potential misuse introduces new security risks [9]. Some studies have investigated malicious RIS configurations that threaten physical-layer security. For example, [10] proposed a threat model where RISs assist eavesdroppers and introduced joint transmit power and RIS configuration optimization. In [11], benign and malicious RIS interactions were studied for MISO wiretap channels, using a game-theoretic max-min secrecy rate optimization under perfect CSI.

Similarly, [12] studied destructive beamforming by adversarial RISs and showed that phase misalignment can degrade secrecy capacity and throughput. More recently, [13] provided a taxonomy of RIS-based attacks (e.g., jamming, eavesdropping, pilot contamination) and their impact in 6G, focusing on control-layer threats and adversarial optimization. Complementing these, [14] proposed an explainable adversarial learning framework to defend against

TABLE I: Comparison of Recent Works on Malicious RIS Behavior

Ref.	Focus	Attack Types	System Assumptions	Main Contributions
[10]	Eavesdropping enhancement using adversarial RIS	RIS-enhanced eavesdropping	Perfect CSI, Eve controls RIS	Introduces RIS as a tool for advanced eavesdropping under idealized assumptions
[11]	Game-theoretic modeling of RIS behavior	Passive jamming by adversarial RIS	Perfect CSI, discrete RIS phases, multiple RIS	Studies RIS interaction under benign and malicious roles using game theory
[12]	Destructive phase-shift design by compromised RIS	SNR degradation via adversarial beamforming	Perfect CSI, passive RIS	Shows effectiveness of malicious RIS beamforming under CSI uncertainty
[13]	Broad taxonomy of control- and signal-level RIS threats	Jamming, pilot contamination, signal leakage, ND-RIS	Partial CSI, passive/active RIS	Comprehensive survey and simulation of multi-modal RIS attack strategies
[14]	ML-based detection of RIS threats during key generation	Adversarial RIS interference and phase manipulation	Partial CSI, adversarial RIS in key generation	Uses explainable ML for malicious RIS detection in key generation
This Work	Hardware-level RIS vulnerabilities	Power-splitting, Element-splitting	Perfect CSI, passive RIS	Models partial hardware compromise of RIS, and quantify its physical-layer impact

malicious RIS behavior during physical-layer key generation. While promising, most works assume adversarial control at the configuration or algorithmic level.

While prior studies highlight RIS security, hardware-level vulnerabilities from physical deployment remain underexplored. Unlike conventional wireless nodes, RISs are installed on exposed surfaces without sensing or authentication, making them prone to tampering, malicious firmware, or element control during manufacturing. These risks are amplified by RIS-specific traits like passive operation and centralized control. Our work focuses on this underexplored attack surface by modeling low-level compromises, specifically, power-splitting and element-splitting strategies, and demonstrating their impact on system performance. The comparison in Table I contextualize our approach relative to existing system-level and algorithm-level RIS threat models. By addressing threats that cannot be mitigated through signal processing alone, this work complements existing research and underscores the need for hardware-aware RIS security frameworks.

Our goal is to emphasize that RIS security must be considered from the hardware design level rather than added later. By highlighting the intersection of hardware security and emerging wireless architectures, this work seeks to spark broader dialogue on the safe and trustworthy deployment of RIS in future wireless networks. In summary, the following are our key contributions in this paper.

- We introduce the concept of *Malicious RIS* (MALRIS), where hardware-compromised RIS components covertly degrade communication or assist eavesdropping, even during passive operation.
- We identify and categorize practical hardware-level threats, including tampering, firmware manipulation, and electromagnetic interference, distinguishing them from traditional signal-level attacks.
- We propose two stealthy attack models, *power-splitting* and *element-splitting*, to represent partial RIS compromise, and analyze their impact on confidentiality, reliability, and availability.
- We evaluate these threats through simulations, showing their significant effect on bit error rate (BER), secrecy capacity, and outage probability, even with limited hardware manipulation.

The rest of the paper is structured as follows: Section II throws light on the hardware-based threats in RIS. Security implications of these attacks and risk assessment are presented in Section III. Impact of MALRIS on the different performance metrics is detailed in Section IV. Finally, the paper concludes in Section V with some future directions.

II. Hardware-based Threats to RIS-assisted communication

RIS introduces unique vulnerabilities due to its reconfigurable hardware and reliance on external control circuits, creating new attack surfaces, especially under physical or electromagnetic access. This section outlines key hardware-level threats to RIS integrity and functionality. We consider a typical RIS-assisted communication scenario where a base station (BS) serves a user equipment (UE) via a RIS managed by an external controller (as in Fig. 1a). An

Eve nearby may exploit hardware vulnerabilities in the RIS or its control path. This highlights the hardware-centric attack surface inherent in RIS deployment, which is examined next.

A. RIS Element Tampering

RIS comprises dense arrays of programmable elements (e.g., FPGAs, PIN diodes, varactors, MEMS switches) to adjust wave phases [1]. Often deployed on exposed structures (Fig. 1b), they are vulnerable to tampering, with MALRIS reflecting compromised signals to the UE. Assuming trusted design and manufacturing, physical attackers can still damage or rewire components, degrading beamforming and secrecy. As RIS evolves toward mmWave/THz 6G, miniaturization increases susceptibility to disruption [15]. Even without physical access, high-power electromagnetic surges can induce faults remotely [16].

B. Malicious Reconfiguration via Control Interface

In addition to element tampering, RIS controllers are vulnerable as they manage phase shifts via commands from a central node over wired or wireless links. Without baseband processing or autonomous control, RIS primarily relies on control signal integrity. As shown in Fig. 1c, the control interface can be a single point of failure if compromised. Here, we assume RIS controllers and elements are compromised during design or manufacturing, with no man-in-the-middle attacks. Malicious actors may exploit the supply chain by embedding hidden circuits or hardware Trojans in RIS components, triggered under specific conditions [17]. FPGA-based RIS can be reconfigured via illegitimate bitstreams [18], while logic bombs may activate under certain RF conditions [19]. Without encrypted control channels, attackers can inject spoofed or replayed commands [20]. These hard-to-detect threats can cause coordination failures in multi-panel setups, making the control interface a key security bottleneck.

C. Firmware Compromise and Backdoor

Beyond element and control interface risks, RIS controllers face firmware-level threats. These controllers manage phase settings but usually lack security monitoring. Firmware may be compromised during manufacturing via logic bombs or hidden routines [21], triggered by RF conditions, timing, or post-configuration [20], altering phases or leaking data. As shown in Fig. 1c, we consider trusted hardware with bitstreams maliciously modified during updates, physically or remotely. Unlike conventional nodes, RIS lacks user-facing software and runtime diagnostics, making firmware tampering highly stealthy. Insecure updates without cryptographic authentication [22] let attackers install persistent malware for surveillance or denial-of-service. RIS's passive design, limited processing, and exposed deployment amplify these risks, making firmware-level threats especially critical.

D. Side-Channel Attacks

Even without direct access, attackers can exploit side-channel leaks, e.g., electromagnetic radiation, power fluctuations [23], or acoustic signals (as in Fig. 1d), from RIS control circuitry, revealing phase settings, beam directions, or user activity. Here, we assume trusted design and manufacturing, but the channel between the RIS controller and the RIS is compromised by man-in-the-middle attacks. An attacker can passively observe side-channel patterns to infer operations, correlate data, or fingerprint RIS panels and firmware, without system contact. These stealthy attacks need no modifications, use off-the-shelf tools, leave no trace, and are hard to defend against due to RIS's passive design.

III. Security Implications and Risk Assessment

Hardware-level attacks on RIS components pose serious risks to wireless system security. This section analyzes these threats using the Confidentiality, Integrity, and Availability (CIA) framework to highlight how they can enable unauthorized access, disrupt services, or degrade performance in real-world deployments.

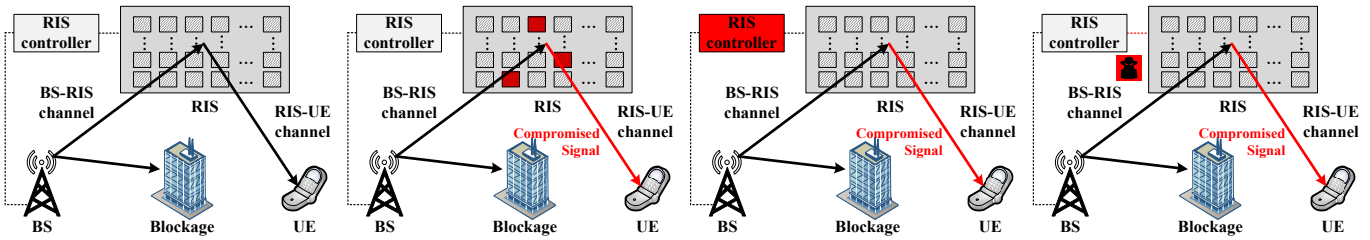


Fig. 1: (a) RIS-assisted network mode (b) RIS element tampering (c) Compromised RIS controller (d) Side-channel attack.

A. Confidentiality Risks

Confidentiality in RIS-assisted networks is especially vulnerable due to their passive nature and reliance on external control. Adversaries can exploit several threats to intercept sensitive information silently: *a) Malicious reconfiguration* can steer signals toward nearby eavesdroppers. *b) Backdoor insertion*, such as a hardware Trojan, during device manufacturing, may leak configuration or channel data. *c) Side-channel leakage* reveals beam directions or user movement through emissions. These attacks bypass traditional cryptographic defenses by targeting the physical layer, enabling undetectable surveillance and undermining the spatial privacy that directional systems are meant to ensure.

B. Integrity Risks

Integrity in RIS-assisted systems is particularly fragile due to their centralized, passive control. Adversaries can manipulate RIS behavior to cause misleading or degraded operations through: *a) element tampering* involves physically altering reflectors to misdirect beams or degrade precision, *b) firmware compromise* by injecting malicious logic to trigger unauthorized phase changes or noise only during operation, and *c) spoofed commands* to mislead the RIS into applying attacker-defined configurations. Such attacks may go undetected yet significantly disrupt communication. In cooperative or high-precision systems, even subtle integrity breaches can cascade, destabilizing the entire transmission process.

C. Availability Risks

Availability attacks on RIS aim to disrupt service or disable components, and are especially dangerous due to RIS's passive design and dependence on precise control. Key threats include: *a) Physical tampering* or EM interference can damage or desynchronize RIS elements, degrading coverage. *c) Unauthorized commands* may overload the system with invalid reconfigurations, disabling functionality. *d) Malicious firmware* updates during operation can "brick" the RIS, with no built-in recovery. *e) Jamming amplification*, where compromised hardware in RIS reflects and enhances jamming signals, extends the attack's reach and impact. These threats can cripple systems, especially in cooperative MIMO or URLLC, while escaping detection due to minimal RIS monitoring. What was designed to enhance performance can become a tool for disruption.

IV. Performance Evaluation and Discussions

This section presents the system model and evaluates the performance of MALRIS. We examine two scenarios where MALRIS redirects signals to Eve: by manipulating a subset of elements or diverting part of the signal power.

A. System Model

We consider a RIS-assisted wireless system where a BS with M antennas serves a single-antenna UE via a RIS with N elements. Perfect CSI for the BS-RIS-UE path is assumed. The RIS is malicious, which means MALRIS operates normally, assisting BS-to-UE communication. However, when a nearby Eve is present, it maliciously redirects part of the BS signal toward Eve.

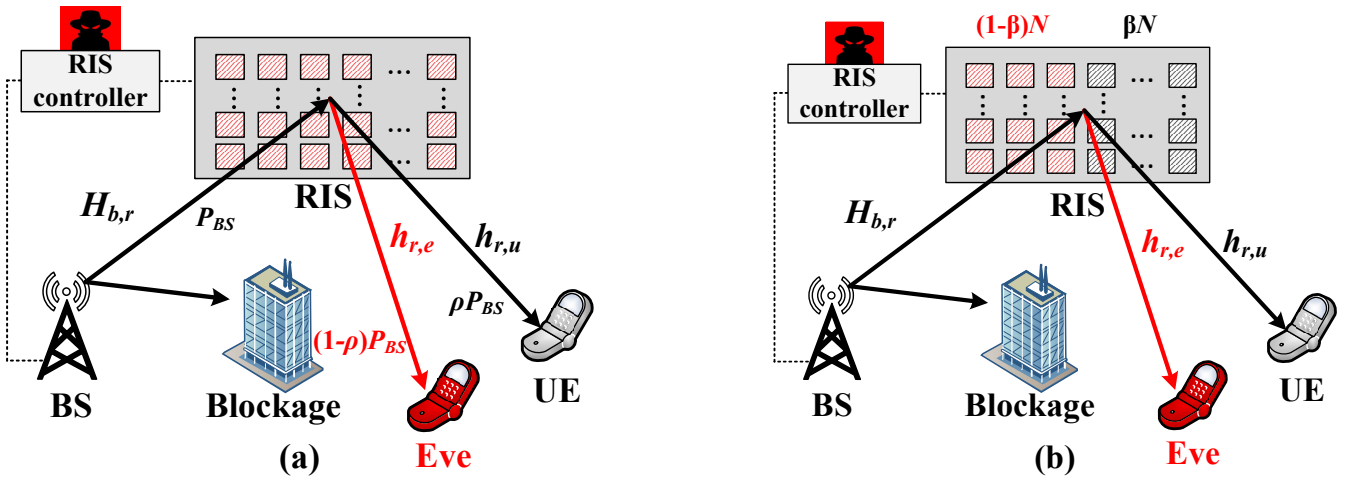


Fig. 2: RIS-assisted communication under MALRIS threats, illustrating signal leakage via malicious reflection strategies: (a) power splitting and (b) element splitting.

1. Channel Modeling

Let the channel from BS to RIS be denoted as $\mathbf{H}_{\text{b,r}} \in \mathbb{C}^{N\times M}$, and the channel from RIS to UE be $\mathbf{h}_{\text{r,u}} \in \mathbb{C}^{1\times N}$. We assume that each channel follows a Rician fading model such that,

$$\mathbf{H}_{\text{b,r}} = \sqrt{PL_{\text{b,r}}}\left(\sqrt{\frac{\kappa}{\kappa+1}}\mathbf{H}_{\text{b,r}}^{\text{LOS}} + \sqrt{\frac{1}{\kappa+1}}\mathbf{H}_{\text{b,r}}^{\text{NLOS}}\right), \tag{1}$$

$$\mathbf{h}_{\text{r,u}} = \sqrt{PL_{\text{r,u}}}\left(\sqrt{\frac{\kappa}{\kappa+1}}\mathbf{h}_{\text{r,u}}^{\text{LOS}} + \sqrt{\frac{1}{\kappa+1}}\mathbf{h}_{\text{r,u}}^{\text{NLOS}}\right), \tag{2}$$

where κ is the Rician factor, and $PL_{\text{b,r}}$ and $PL_{\text{r,u}}$ denote large-scale path loss terms modeled as $PL = d^{-\alpha}$, with d being the distance between two network entities and α is the path-loss exponent. Similarly, the channel from RIS to Eve is denoted by $\mathbf{h}_{\text{r,e}} \in \mathbb{C}^{1\times N}$, also modeled via Rician fading and subject to large-scale fading $PL_{\text{r,e}}$.

B. Signal Model Without MALRIS

Let the BS transmit signal s with power P_{BS} using beamforming vector $\mathbf{x} \in \mathbb{C}^{M\times 1}$ satisfying $\|\mathbf{x}\|^2 = P_{\text{BS}}$. The beamforming vector $\mathbf{x}$ is designed using Maximum Ratio Transmission (MRT) based on the effective cascaded channel. Assuming perfect CSI of the BS–RIS–UE link, the optimal beamforming direction is given by:

$$\mathbf{x} = \sqrt{P_{\text{BS}}} \cdot \frac{\left(\mathbf{H}_{\text{b,r}}^{\dagger}\boldsymbol{\Theta}^{H}\mathbf{h}_{\text{r,u}}^{H}\right)}{\left\|\mathbf{H}_{\text{b,r}}^{\dagger}\boldsymbol{\Theta}^{H}\mathbf{h}_{\text{r,u}}^{H}\right\|}, \tag{3}$$

where $\dagger$ denotes the Hermitian (conjugate transpose), and $\|\cdot\|$ is Frobenius norm. This ensures the BS transmits along the strongest direction of the cascaded channel $\mathbf{h}_{\text{r,u}}\boldsymbol{\Theta}\mathbf{H}_{\text{b,r}}$, maximizing the received power at the UE. Let $\boldsymbol{\Theta} = \text{diag}(e^{j\theta_1}, ..., e^{j\theta_N})$ be the RIS phase shift matrix optimized using perfect CSI of the BS-RIS-UE link. The received signal at the UE is given by:

$$y_{\text{u}} = \mathbf{h}_{\text{r,u}}\boldsymbol{\Theta}\mathbf{H}_{\text{b,r}}\mathbf{x}s + n_{\text{u}}, \tag{4}$$

where $n_{\text{u}} \sim \mathcal{CN}(0, \sigma_u^2)$ is the additive white Gaussian noise. $s \sim \mathcal{CN}(0, 1)$ is a transmitted symbol represented by zero-mean circularly symmetric complex Gaussian random variables with unit power satisfying $\mathbb{E}[|s|^2] = 1$. The effective end-to-end channel gain is: $h_{\text{eff}}^{(u)} = \mathbf{h}_{\text{r,u}}\boldsymbol{\Theta}\mathbf{H}_{\text{b,r}}\mathbf{x}$, and the resulting SNR (γ_{u}) at UE is:

$$\gamma_{\text{u}} = \frac{|h_{\text{eff}}^{(u)}|^2}{\sigma_u^2} = \frac{|\mathbf{h}_{\text{r,u}}\boldsymbol{\Theta}\mathbf{H}_{\text{b,r}}\mathbf{x}|^2}{\sigma_u^2}. \tag{5}$$

C. Signal Model During MALRIS Behavior

We assume that the MALRIS starts the security attack after $T/2$ time slots (T is the total number of slots), whereby it splits its reflection resources between the UE and Eve. In this paper, we consider two scenarios: power splitting and RIS element splitting.

(a) Power-based Splitting: We studied a possible attack whereby MALRIS splits incoming signal power such that ρP_{BS} is directed toward the UE and $(1-\rho)P_{\text{BS}}$ is reflected toward the Eve, with $\rho \in [0,1]$ is the power splitting factor. Specifically, signal power can be manipulated by attack vectors such as malicious reconfiguration via control interface. In this scenario, effective channel gain and SNR (γ_{u}) at the UE are given as

$$h_{\text{eff}}^{(u)} = \sqrt{\rho} \cdot \mathbf{h}_{\text{r,u}} \boldsymbol{\Theta} \mathbf{H}_{\text{b,r}} \mathbf{x}, \tag{6}$$

$$\gamma_{\text{u}} = \frac{|h_{\text{eff}}^{(u)}|^2}{\sigma_u^2}. \tag{7}$$

For Eve, the received signal is modeled statistically as $y_{\text{e}} = \sqrt{1-\rho} \cdot z + n_{\text{e}}$, with the resulting SNR (γ_{e}) as

$$\gamma_{\text{e}} = \frac{(1-\rho)|z|^2}{\sigma_e^2}. \tag{8}$$

Here, z is an unknown leakage signal and γ_{e} is approximated and simulated empirically.

(b) Element-based Splitting: In addition to power-based splitting, MALRIS can manipulate a subset of RIS elements to facilitate spoofing or side-channel attacks. Let $\beta \in [0,1]$ represent the fraction of RIS elements allocated to the UE. The remaining fraction $(1-\beta)$ is used by MALRIS to reflect the signal to Eve. The MALRIS phase matrix can be partitioned accordingly, and the effective channel for UE in this scenario is computed as:

$$h_{\text{eff}}^{(u)} = \mathbf{h}_{\text{r,u}}^{(\beta)} \boldsymbol{\Theta}_{\text{u}} \mathbf{H}_{\text{b,r}}^{(\beta)} \mathbf{x}, \tag{9}$$

$$\gamma_{\text{u}} = \frac{|h_{\text{eff}}^{(u)}|^2}{\sigma_u^2}. \tag{10}$$

Here, $\mathbf{H}_{\text{b,r}}^{(\beta)}$ is the subset of $\mathbf{H}_{\text{b,r}}$ matrix, representing the channels for RIS-UE communication. Since the CSI of the RIS-Eve link is unknown, the effective signal at Eve is modeled as $y_{\text{e}} = z + n_{\text{e}}$, and the resultant SNR at Eve γ_{e} is

$$\gamma_{\text{e}} = \frac{|z|^2}{\sigma_e^2}, \tag{11}$$

where z is a random leakage signal dependent on $(1-\beta)N$ RIS elements, and $n_{\text{e}} \sim \mathcal{CN}(0, \sigma_e^2)$ is noise with power σ_e^2.

D. Performance Evaluation

We evaluate throughput, secrecy capacity, secrecy outage probability, and (BER) as performance metrics for the proposed model under two scenarios: with and without a compromised RIS. For the first half of the simulation time, i.e., during $T/2$ time slots (where T is the total simulation time), the RIS operates normally. After $T/2$, the RIS becomes malicious and begins splitting either its antenna elements or the transmitted power between the legitimate user and the Eve. Throughput is computed in terms of Shannon capacity as

$$C = \log_2(1+\gamma), \quad \text{[bps/Hz]}, \tag{12}$$

and secrecy capacity, defined as the maximum rate at which secure communication can be achieved [24], given as:

$$C_s = \max\left(0, \log_2(1+\gamma_{\text{u}}) - \log_2(1+\gamma_{\text{e}})\right). \tag{13}$$

Moreover, we computed secrecy outage probability (P_{out}), which is the probability that the secrecy capacity is less than the target secrecy rate (R_s), given as $P_{\text{out}} = \Pr(C_s < R_s)$. We have estimated P_{out} via Monte Carlo simulation for performance evaluation over multiple channel realizations as:

$$P_{\text{out}} \approx \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} \mathbb{1} \cdot (C_s^{(i)} < R_s), \tag{14}$$

TABLE II: Simulation Parameters

Parameter	Value	Parameter	Value
BS location	(0,0)	RIS location	(50,20)
UE location	(75,0)	Eve location	(50,-20)
BS power (P_{BS}) (db)	10, 20	RIS elements (N)	32, 64
BS antenna (M)	4	Noise power σ_u^2, σ_e^2	10^{-7}
Rician factor (κ)	5	Path loss exponent (α)	3
Simulation time (T)	50	Target secrecy rate (R_s)	1 bps/Hz

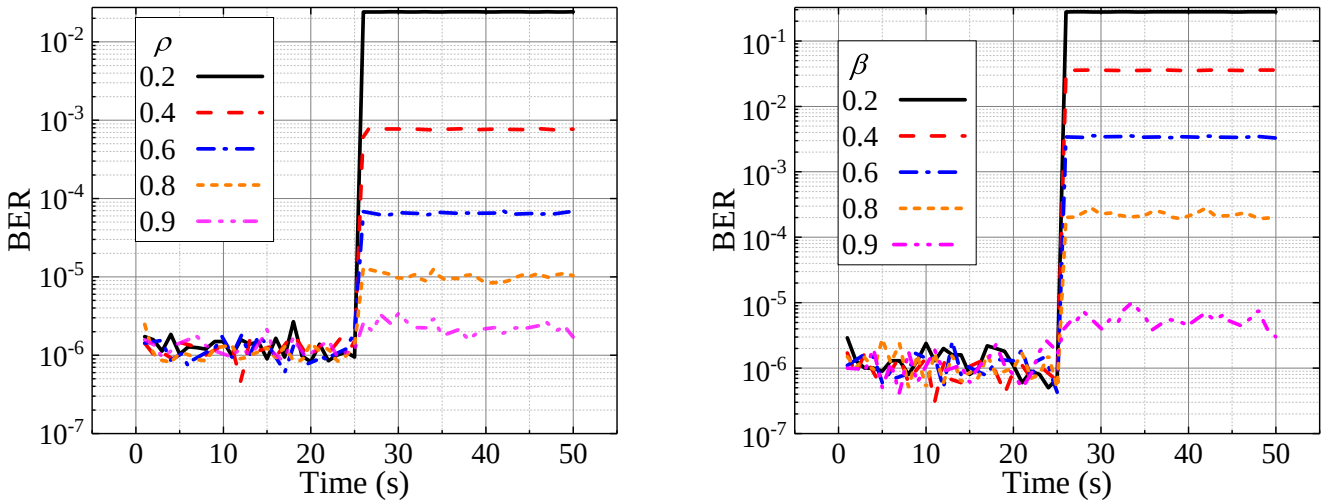
where $\mathbb{1}$ is the indicator function.

Lastly, to compute the BER at UE, we modulated random binary data using QPSK (where two bits are mapped per symbol), transmitted through the channel, and demodulated at the receiver. The BER is computed as the ratio of incorrectly decoded bits to the total number of transmitted bits.

The security vulnerabilities when Eve manipulates some of the antenna elements and power splitting for malicious activities are presented in terms of BER in Fig. 3. General parameter settings for these results are summarized in Table II. As illustrated in Figs. 3a and 3b, the BER increases sharply when a malicious attack is launched by MALRIS. Intuitively, for lower values of ρ, less power is allocated for the UE, resulting in degraded BER due to weak signal strength, as shown in Fig. 3a. In contrast, higher ρ values result in reflecting the signal to the UE with more power, improving the BER and resulting SNR. Fig. 3b further demonstrates the impact of the antenna element splitting on BER performance during a MALRIS attack. Lower β values result in fewer RIS elements allocation for the UE and degraded BER performance due to weaker signal strength. Comparing the two figures, it is evident that antenna splitting results in worse BER performance than power splitting. This degradation is primarily attributed to the significant reduction in the RIS gain available to the UE during antenna splitting.

Fig. 4 shows the throughput of the UE and Eve as functions of ρ and β. In Fig. 4a, throughput is plotted against ρ for different P_{BS} values. When ρ is small, most of the power is directed to Eve, resulting in high Eve throughput and low UE throughput. As ρ increases, UE throughput improves while Eve's decreases. Higher P_{BS} leads to greater throughput in both cases owing to the higher signal strength at the receiving entities. However, at mid-range ρ values, UE and Eve throughput become comparable, raising security concerns. Fig. 4b illustrates throughput versus the antenna splitting factor β. Increasing β allocates more elements to the UE, enhancing its throughput. Hereby, as Eve is farther from RIS compared to UE, its performance is mostly lower, when β increases as the RIS gain becomes minimal. The RIS gain is noticeable for the UE case, whereby the difference between $N = 32$ and $N = 64$ increases as the value of β increases.

Fig. 5 illustrates the secrecy outage probability (P_{out}) and secrecy capacity (C_s) of the UE under power and antenna splitting scenarios. In the power splitting case (Fig. 5a), P_{out} drops sharply after $\rho = 0.4$ and eventually



(a) (b)

Fig. 3: BER performance of MALRIS-assisted network under (a) power splitting and (b) antenna element-splitting attacks.

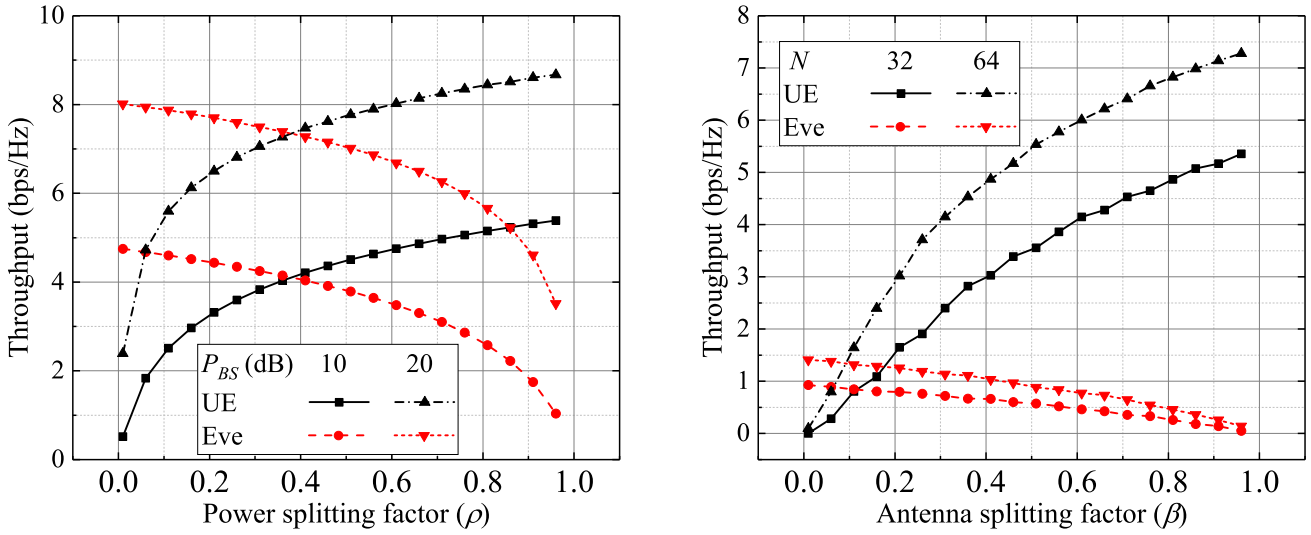


Fig. 4: Throughput of the MALRIS-assisted network under (a) power-splitting and (b) antenna element-splitting attacks.

reaches zero, indicating a significant improvement in UE throughput relative to Eve, as also seen in Fig. 4a. This results in a corresponding rise in C_s. Notably, the difference in P_{out} and C_s between 10 dB and 20 dB P_{BS} is minimal, as the power allocation ratio between UE and Eve remains unchanged. In contrast, Fig. 5b (antenna splitting case) shows a widening gap in both P_{out} and C_s as β increases, primarily due to enhanced RIS gain. At lower N, the secrecy outage is higher, but it decreases rapidly with increasing β, since fewer RIS elements are reflecting toward Eve and its signal weakens with distance. The secrecy capacity improves accordingly, with a growing performance gap between $N = 32$ and $N = 64$, consistent with the trend observed in Fig. 4b for the throughput of UE.

V. Conclusion and Future Directions

RIS offers transformative potential for wireless communication through programmable radio environment control. However, its architecture introduces hardware-level vulnerabilities beyond traditional security models. This paper identified RIS-specific hardware threats, ranging from physical tampering and malicious reconfiguration to side-channel attacks, and highlighted their impact on confidentiality, integrity, and availability. These stealthy, trace-resistant threats underscore the urgent need for end-to-end security from design and manufacturing to deployment.

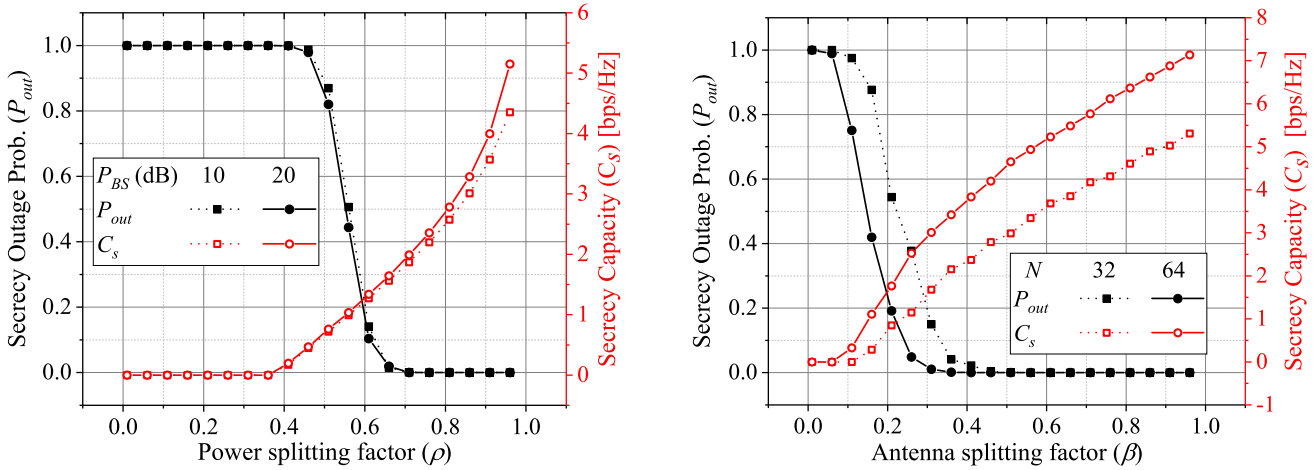
Future work will explore broader hardware-level attacks and practical defenses, including physical protection, secure reconfiguration, and trusted firmware. We also aim to investigate AI for real-time threat mitigation and RIS integration in multi-agent systems, where scalable, robust security is critical.

Acknowledgments

This research was supported by the Sungkyunkwan University and the BK21 FOUR(Graduate School Innovation) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2024-00397216), the German Federal Ministry of Education and Research (BMBF) within the projects Open6GHub under grant number {16KISK004}, and under grant number {03FHP106}, as part of the "Career@BI" project within the FH Personal program.

References

[1] E. Basar, G. C. Alexandropoulos, Y. Liu, Q. Wu, S. Jin, C. Yuen, O. A. Dobre, and R. Schober, "Reconfigurable Intelligent Surfaces for 6G: Emerging Hardware Architectures, Applications, and Open Challenges," *IEEE Vehicular Technology Magazine*, vol. 19, no. 3, pp. 27–47, 2024.



(a) (b)

Fig. 5: Secrecy outage probability and capacity of the MALRIS-assisted network under (a) power-splitting and (b) antenna element-splitting attacks.

[2] E. Björnson, O. Özdogan, and E. G. Larsson, "Reconfigurable Intelligent Surfaces: Three Myths and Two Critical Questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.

[3] W. Jiang and F.-L. Luo, *Intelligent Reflecting Surface-Aided Communications for 6G*. IEEE-Wiley, 2023, pp. 295–362.

[4] Q. Wu, B. Zheng, C. You, L. Zhu, K. Shen, X. Shao, W. Mei, B. Di, H. Zhang, E. Basar, L. Song, M. Di Renzo, Z.-Q. Luo, and R. Zhang, "Intelligent Surfaces Empowered Wireless Network: Recent Advances and the Road to 6G," *Proceedings of the IEEE*, vol. 112, no. 7, pp. 724–763, 2024.

[5] Z. Li, O. A. Topal, Ö. T. Demir, E. Björnson, and C. Cavdar, "mmWave Coverage Extension Using Reconfigurable Intelligent Surfaces in Indoor Dense Spaces," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 5805–5810.

[6] R. K. Fotock, A. Zappone, and M. D. Renzo, "Energy Efficiency Optimization in RIS-Aided Wireless Networks: Active Versus Nearly-Passive RIS With Global Reflection Constraints," *IEEE Transactions on Communications*, vol. 72, no. 1, pp. 257–272, 2024.

[7] H. Shakhatreh, A. Sawalmeh, K. F. Hayajneh, S. Abdel-Razeq, W. Malkawi, and A. Al-Fuqaha, "A Systematic Review of Interference Mitigation Techniques in Current and Future UAV-Assisted Wireless Networks," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2815–2846, 2024.

[8] M. Guo, Z. Lin, R. Ma, K. An, D. Li, N. Al-Dhahir, and J. Wang, "Inspiring Physical Layer Security With RIS: Principles, Applications, and Challenges," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 2903–2925, 2024.

[9] B. Lyu, D. T. Hoang, S. Gong, D. Niyato, and D. I. Kim, "IRS-Based Wireless Jamming Attacks: When Jammers Can Attack Without Power," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1663–1667, 2020.

[10] G. C. Alexandropoulos, K. D. Katsanos, M. Wen, and D. B. Da Costa, "Counteracting Eavesdropper Attacks Through Reconfigurable Intelligent Surfaces: A New Threat Model and Secrecy Rate Optimization," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1285–1302, 2023.

[11] Y. Gao, S. Rezvani, P.-H. Lin, and E. A. Jorswieck, "Benign and Malicious Reconfigurable Intelligent Surfaces in MISO Wiretap Channels," in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2024, pp. 541–545.

[12] S. Rivetti, Ö. T. Demir, E. Björnson, and M. Skoglund, "Malicious Reconfigurable Intelligent Surfaces: How Impactful Can Destructive Beamforming be?" *IEEE Wireless Communications Letters*, vol. 13, no. 7, pp. 1918–1922, 2024.

[13] W. Khalid, T. V. Chien, W. U. Khan, Z. Kaleem, Y. B. Zikria, T. Kim, and H. Yu, "Malicious Reconfigurable Intelligent Surfaces: Security Threats in 6G Networks," *IEEE Internet of Things Journal*, pp. 1–1, 2025.

[14] Z. Wei, W. Hu, J. Zhang, W. Guo, and J. A. McCann, "Explainable Adversarial Learning Framework on Physical Layer Key Generation Combating Malicious Reconfigurable Intelligent Surface," *IEEE Transactions on Wireless Communications*, vol. 24, no. 4, pp. 3529–3545, 2025.

[15] W. Jiang, Q. Zhou, J. He, M. A. Habibi, S. Melnyk, M. El-Absi, B. Han, M. D. Renzo, H. D. Schotten, F.-L. Luo, T. S. El-Bawab, M. Juntti, M. Debbah, and V. C. M. Leung, "Terahertz Communications and Sensing for 6G and Beyond: A Comprehensive Review," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 4, pp. 2326–2381, 2024.

[16] Y. Tian, Y. Song, Y. Li, S. Luan, S. Xie, Y. Yang, and Y. Li, "Microwave Photon Receiving Antenna: New Concept and UWB Applications," *IEEE Transactions on Antennas and Propagation*, vol. 73, no. 3, pp. 1383–1393, 2025.

[17] S. Akter, K. Khalil, and M. Bayoumi, "A Survey on Hardware Security: Current Trends and Challenges," *IEEE Access*, vol. 11, pp. 77 543–77 565, 2023.

[18] Q. A. Ahmed, T. Wiersema, and M. Platzner, "Malicious Routing: Circumventing Bitstream-level Verification for FPGAs," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2021, pp. 1490–1495.

[19] Y. Jin, D. Maliuk, and Y. Makris, *Hardware Trojan Detection in Analog/RF Integrated Circuits*. Cham: Springer International Publishing, 2016, pp. 241–268.

[20] Q. A. Ahmed, T. Wiersema, and M. Platzner, "Post-configuration Activation of Hardware Trojans in FPGAs," *Journal of Hardware and Systems Security*, vol. 9, no. 10, pp. 2509–3436, 2024.
[21] K. S. Subramani, N. Helal, A. Antonopoulos, A. Nosratinia, and Y. Makris, "Amplitude-Modulating Analog/RF Hardware Trojans in Wireless Networks: Risks and Remedies," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3497–3510, 2020.
[22] C. Krieg, C. Wolf, and A. Jantsch, "Malicious LUT: A stealthy FPGA Trojan injected and triggered by the design flow," in *2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2016, pp. 1–8.
[23] F. Schellenberg, D. R. Gnad, A. Moradi, and M. B. Tahoori, "An inside job: Remote power analysis attacks on FPGAs," in *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2018, pp. 1111–1116.
[24] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *2006 IEEE International Symposium on Information Theory*, 2006, pp. 356–360.